DOCUMENT RESUME

ED 121 102 FL 007 589

AUTHOR Bebout, Linda
TITLE An Error Analysis: Comparing the Ability of Learners of English as a First and as a Second Language to Extract Information from Written Material.
PUB DATE Mar 75
NOTE 15p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Cloze Procedure; *Comparative Testing; Contrastive Linguistics; English; *English (Second Language); Error Patterns; Language Instruction; Language Research; *Language Tests; Native Speakers; Second Language Learning; Spanish Speaking
IDENTIFIERS *Error Analysis

ABSTRACT

The major purposes of the study reported on in this paper were as follows: (1) to compare the errors made by advanced learners of English as a native language (American nine- to eleven-year-olds) and as a foreign language (native Spanish-speaking adults); (2) to investigate the usefulness of a modified cloze test in obtaining errors for profitable analysis; and (3) to aid teachers and researchers in evaluating the potential of error analysis as a tool by providing a demonstration of what it can reveal about language learning. Given a test consisting of a series of highly-contextualized cloze-type items, the subjects were asked to supply a word for each blank which would satisfactorily complete the sentence. Those responses judged by adult native English speakers to be errors were classified by the experimenter into six (non-exclusive) categories, entitled context alteration, non sequiturs, grammatical errors, lexical confusions, uninterpretable, and other (more minor). The individual error categories and the pattern of error contributions by the two subject groups are discussed, although the errors contributed to the categories by the two groups showed an overall similarity both in terms of quantity and quality. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121102

An Error Analysis: Comparing the Ability of Learners of English As a First and as a Second Language to Extract Information from Written Material[1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL007589

Linda Bebout
St. Catharines, Ontario, Canada

Interest in the use of error analysis as a tool for the study of language learning has been growing rapidly in the last few years. (See, for example, Sampson & Richards 1973, Buteau 1970, Dušková 1969, George 1972, Taylor 1975.) The question of the relationship of first and second language learning has also begun to be discussed and investigated more seriously, often by means of error an ysis (e.g. Dulay & Burt 1972, Ravem 1970). Another tool which is proving to be increasingly useful in the investigation of language learning is that of cloze technique (e.g. Darnell 1968, Oller 1973, Oller & Inal 1971). The present study (Bebout 1974) combines a type of cloze test with error analysis to explore and compare the performance of a group of second language learners with that of a group of first language learners on a task involving the extraction and use of information from written material.

A traditional cloze test consists of a passage of written material which has been "mutilated" by deleting every nth word (or random words). The test is administered by asking the subject ro fill in the blanks with the words which he thinks were deleted. Although it was originally devised as a means of determining the readability of a text (Taylor 1953), the cloze test is also proving to be useful as a global, integrative test of the language ability of the person who takes it. A modified version of the classical cloze test was used in this study. The five modifications were as follows: (i) one- or two-sentence items were used rather than whole passages; (ii) the items were written with blanks, rather than having words deleted from original complete sentences; (iii) there was only one blank per item; (iv) the subjects were instructed to respond with any word which they thought would "make sense" in the blank; and (v) the items were written with the intent of providing a fairly constraining context for the blank, i.e., in order for his response to make sense, the subject had to pick up information from the item and use it. There

were no sentences like

This is a _____.

or

_____ went away.

Instead, the following is a typical item:

I really like the color of your _____. Where did you buy him?

These modifications resulted in items which closely resembled traditional completion test items, except that there was obviously no single "correct" response for any item (contrary to what is the case in most language completion tests).

By constructing the items as described, it was hoped that a profitable, informative error analysis of the responses could be made. The major tenet of error analysis is that the learner's errors can help to reveal how he is going about learning the language (Corder 1967). Thus, a quantitative and qualitative comparison of the errors made by the two subject groups could be of value in comparing their approaches to language learning.

One subject group consisted of 61 nine-to-eleven-year-old monolingual American school children and the other of 84 native Spanish-speaking Mexican adults who were advanced students of English as a foreign language. Both groups were thus considered to be advanced learners of English, either as a first or a second language. No selection procedures were used; in both the American and the Mexican settings, all of the students in several classes were tested. The test was in English, although Spanish instructions were also provided for the adults. The linguistic material used in the test items was controlled by reference to relevant textbooks and word lists to ensure that it was in general neither beyond the reading level of the children nor beyond the English level of the adults. (To encourage a maximum number

of responses, the children were told that it was permissible to use words whose spelling they were unsure of and the adults were told that Spanish responses [later translated by the experimenter] were allowed).

The subjects' responses were rated by six native-English-speaking adults as to their acceptability as adult, native-speaker English utterances. (See Appendix A for rating scale used.) The average of these ratings was used to determine which responses would be considered as 'errors' and used in the error analysis. A response which received ratings other than "completely acceptable" from more than one half of the raters was considered to be an error for purposes of analysis. There was no significant difference between the two subject groups in the percentages of their responses which were defined as errors. That is, both groups had the same number of errors per responses given. The responses which had been rated as unacceptable were then classified by the experimenter into one or more of the following categories:

I. Response is made more acceptable by altering context

   A. Deletion
      1. Before blank
      2. After bland

   B. Replacement
      1. Before blank
      2. After blank

   C. Addition
      1. Before blank
      2. After blank

   D. Other

II. Non-sequiturs

   A. An explicit connective relationship was ignored or misused

   B. An implicit connective relationship was ignored or misused

   C. The subject's reasoning was erroneous or he failed to take into account a major segment of the item

III. Grammatical errors

A. Nouns

B. Verbs
1. Number
2. Tense/finite vs. non-finite

C. Prepositions
1. Wrong one
2. Omitted
3. Extra

D. Other
1. Incorrect part of speech
2. Inappropriate or non-existent form
3. (1) + semantically infeasible

IV. Lexical errors

A. Within English
1. Native speakers
2. Non-native speakers

B. Spanish/English

V. Uninterpretable

A. No source discernible

B. Repetition of part of context

C. Associative relationship with something in context

D. Misunderstanding of one aspect of instructions

VI. Other (more minor)

A. Awkward style

B. (More) acceptable if external contextual situation supplied by reader

C. Unexpected, or describes improbable situation

D. Shows lack of knowledge of the world.

The following items and responses are representative of those placed in the first category given in the outline above. (The letter and number in front of the item, e.g. A2, indicates the particular subcategory

illustrated by the example. The numbers in parentheses after the responses, e.g. (1-5), indicate the number of non-native speakers and native speakers, respectively, who gave that response.)

(A2) Did you _____ for the tickets, or should I?
buy (1-5) (Correction: delete <u>for</u>)

(B2) How can I _____ you if I don't know the question?
solve (1-1) (replace <u>you</u> with <u>it</u> or a noun)

As can be seen from the table in Appendix B, the two subject groups contributed equivalent proportions of responses to this error category. The major hypothesis for the source of Type I errors was that the subject misread or failed to understand one or two words in the context sentence. Neither the presumed superior familiarity with the language of the native speakers nor the age and reading experience of the adults seemed to give one group an advantage over the other on this type of error.

The following items were among those placed in Category II:

(C) --Do you think this coffee has too much sugar in it?
--No, I like very _____ coffee.
hot (0-1), strong (1-4)

(B) If you think the room is too _____, take off your sweater.
cold (2-3)

There were significantly more native-speaker responses in this category. Two equally-plausible sources were hypothesized for Type II errors: (1) that the subject failed to correctly interpret a connective relationship between two parts of the item, or (2) that he didn't take into account some large segment of the item. Since careless reading habits and an immature understanding of connectives and implied connective relationships could lead to Type II errors, it is not surprising (at least post facto) that the children gave more responses which fell into this category.

Category III:

(B2) --Did Karen come to your party?
--No, I forgot to _____ her.
invited (1-1)

(A) He was the only child at the party. All the others were _____.
man (1-3)

As can be seen from the above examples, the grammatical errors referred to in Category III are not of the prescriptive type (like violation of the who/whom rules), nor are they even patterns acceptable in nonstandard dialects (like he don't). For this reason, I find it surprising that the native speakers were so well-represented in this category, with responses qualitatively indistinguishable from those of the non-native speakers. However, the quantitative predominance of non-native-speaker errors which did indeed occur would be expected.

Category IV errors are exemplified by these two items:

(A1,2) Could you help Tommy get dressed? He's too young to _____ his own shirt.
wear (4-4) (Correction: substitute put on)

(A1) Bring me a stick about two inches thick and three feet _____.
short (0-1), deep (0-1), wide (1-6) (substitute long)

The qualitative-quantitative relationship pattern is much the same in this category as in the previous one: similarities result from the fact that both native-speaking children and non-native-speaking adults are prone to having incomplete or inaccurate knowledge of the semantic features and usage of a lexical item. However, a quantitative difference, with proportionately more non-native-speaker errors, could be expected because the bilinguals also have the confounding factor of their native vocabulary (with this factor being accentuated by the relatively close historical relationship between the Spanish and English languages).

The responses in the fifth category, as exemplified below, were not particularly interesting for the type of error analysis made here, where the subjects were no longer available to the analyst at the time the analysis was done. However, the fact that this category exists suggests that face-to-face interviews are a desirable component of error analysis.

(A) Was David telling the truth, or was he _____?
lots (0-1)

(C) It's very interesting to learn how fish _____.
caviar (1-0)

These two items were among those placed in the last error category:

(B) Did you buy them a _____ tablecloth, or a rectangular one?
fancy (0-1)

(D) If you don't want to get fat, you should eat _____ for dessert instead of cake.
paper (1-0)

The predominance of native-speaker responses in this category, which consisted of errors which were rated by the native speakers as being almost acceptable, might suggest that when a second-language learner makes an error, it tends to be a major one. However, I think there are better explanations in this case, relating to the native speakers' immaturity and sheer lack of experience in the world.

As I already mentioned, the error-to-response ratios for the two subject groups were not significantly different. Even when the six error categories are considered individually (see above, pp. 5-7), there were two where there was no significant difference between the two subject groups (I, V), two where the difference went in one direction (II,VI), and two where it went in the opposite direction (III,IV). I want it to be clear that I do not think this proves that these two subject groups--that is, learners of English as a first language and learners of English as a second language--have the same command of English. However, I do think that this result points out some similarities between the two groups which seem (at least for these subjects and this test) to have outweighed the differences. (And perhaps it was also the case that for each group, many of the presumed advantages for that group relative to the other group were cancelled out by equivalent disadvantages.) Also, many obvious potential differences were neutralized by the test format: aural presentation of the items with

oral responses would probably have produced very different results. However, the essential similarity between the two groups, their role as language learners, remains an important factor. Both are attempting (whether consciously or not) to attain the same skill in processing language as that possessed by the adult native speaker, and both have not yet attained it. In a test format which requires them to gather together information from the printed word and to use it to produce an appropriate language segment, they make errors (i.e., errors relative to adult native standards) which are often similar and sometimes identical. Native adults also show variations in their ability to use the redundancy in language to extract information from linguistic input, but growth in this ability is one of the best signs of progress toward linguistic maturity.

The skill in context use required by the test used in this study has several interlocking components: the reader-listener must (a) use code-specific grammatical cues,such as verbal affixes which signal the number of an associated noun, (b) use code-specific syntactic/semantic cues, such as verbs which admit only animate subjects or do not take objects, (c) utilize intralanguage dependencies, such as relative pronouns which signal a human antecedent, (d) understnad and respect the nuances and restrictions of connective words, phrases, and other devices, and (e) be adept at revising his hypotheses about the input when pragmatic inconsistencies appear. The use of short, highly contextualized language sequences containing a blank to be filled by the subject (such as the test items in this study) seems to be a valuable means of investigating the learner's abilities in the above areas and thus in a large and important part of the total language acquisition experience.

More specifically, the following factors could be cited as contributing to the value of the modified cloze test: (a) the written mode of presentation maximizes the learner's chances of extracting relevant information by

minimizing time and memory limitations; (b) the large number of cues (relative to what would be found in a random language sample) both gives the subject a better opportunity to produce an acceptable response if he is capable of doing so and gives the analyst a better opportunity to pinpoint the factor(s) leading to the production of an unacceptable response; (c) the brevity of the semantically-related segments (one or two sentences long) and the presence of only one blank per item simplify both the subject's and the analyst's tasks by avoiding widely-separated dependencies and the possibility of past errors in comprehension or production influencing future responses, and (d) the comprehension and the production abilities of the subject are tested simultaneously and integratively.

With respect to comprehension and production, it should be noted that not enough is known about their relationship in either the language learner or the skilled language user to merit extending the findings of this study beyond the production/comprehension patterning of the test format used here. While this format does have real-life analogues in the attempt to comprehend a new or a muffled, garbled, or illegible word encountered in context, it cannot be assumed that these subjects would produce similar errors if asked to write original sentences or if given a simple comprehension test with unmutilated sentences. It would also be interesting to know whether subjects would be capable of indicating that they were aware that some of their responses were unacceptable, even though they were unable to provide a more acceptable response; that is, whether their comprehension was in advance of their production.

Footnote

[1]This article is based on a dissertation submitted to Cornell University, Department of Linguistics, in partial fulfillment of the requirements for a doctoral degree. The research was supported by a National Science Foundation Traineeship and a Sigma Xi Grant-in-Aid of Research.

Appendix A

Guidelines for Acceptability Ratings

1. Completely acceptable and "normal"

2. Conditionally acceptable, e.g.
   a. necessary to invent a context outside of sentence
   b. doesn't conflict with anything in context, but is unexpected
   c. takes minor liberties with English grammar, syntax, etc.
   d. shows a lack of knowledge of the world

3. Grammatically unacceptable, but semantically acceptable

4. Unacceptable under any conditions

Appendix B

Proportions of Responses
Contributed to the Error Categories
by the Subject Groups

Error Type	Significance (χ^2, df=1)	Standardized[1] Non-native (adults)		% of Errors Native (9-11 yrs.)		Total Number of Errors by Category[2]
I	$p > .05$	50%	=	50%	*	347
II	$p < .01$	31%	<	69%		240
III	$p < .001$	68%	>	32%		207
IV	$p < .001$	73%	>	23%		126
V	$p > .05$	45%	<	55%	*	479
VI	$p < .05$	40%	<	60%		446

*Nonsignificant differences

[1]Standardization based on the proportion of errors which was contributed to the total by each subject group. E.g., the non-native speakers contributed the same number of Category I errors per all errors as did the native speakers, hence the standardized ratio is 1/1 (50%/50%). However, the non-native speakers contributed only about half as many Category II errors per total errors as did the native speakers, hence the standardized ratio is approximately 1/2 (31%/69%). In other words:

if: $$S = \frac{\%\text{ of errors contributed to Category X by non-native speakers}}{\%\text{ of errors contributed to total errors by non-native speakers}}$$

$$E = \frac{\%\text{ of errors contributed to Category X by native speakers}}{\%\text{ of errors contributed to total errors by non-native speakers}}$$

then standardized percentages are:

for non-native speakers: $\frac{S}{S+E}$ for native speakers: $\frac{E}{S+E}$

[2]Categories are <u>not</u> exclusive; total errors: 1224 (= 21.3% of total responses).

References

Bebout, Linda 1974. An error analysis: comparing the ability of learners of English as a first and as a second language to extract information from written material. Unpublished Ph.D. dissertation, Cornell University.

Buteau, Magdelhayne F. 1970. Students' errors and the learning of French as a second language: a pilot study. International Review of Applied Linguistics 8. 133-145.

Darnell, D.K. 1968. The development of an English language proficiency test of foreign students using a Clozentropy procedure. Bethesda, Md.: ERIC Document Reproduction Service, ERIC ED 024 039.

Dulay, Heidi C., & Burt, Marina K. 1972. Goofing: an indicator of children's second language learning strategies. Language Learning 22.235-252.

Dušková, Libuše. 1969. On sources of errors in foreign language learning. International Review of Applied Linguistics 7.11-36.

George, H.V. 1972. Common Errors in Language Learning. Rowley, Mass.: Newbury House.

Oller, John W. 1973. Cloze tests of second language proficiency and what they measure. Language Learning 23. 105-118.

Oller, John W., & Inal, Nevin. 1971. A cloze test of English prepositions. TESOL Quarterly 5.315-326.

Ravem, Roar. 1970. The development of wh-questions in first and second language learners. Bethesda, Md.: ERIC Document Reproduction Service, ERIC ED 054 654.

Sampson, Gloria, & Richards, Jack. 1973. Learner language systems. Language Sciences 26.18-25.

Taylor, Barry P. 1975. The use of overgeneralization and transfer learning systems by elementary and intermediate students of English as a second language. Paper presented at the 9th convention of the Teachers of English to Speakers of Other Languages, Los Angeles, Calif., March, 1975.

Taylor, Wilson L. 1953. Cloze procedure: a new tool for measuring readability. Journalism Quarterly 30.415-433.